

SIRA Brown Bag Talk



Title: Apertus: Democratizing Open and Compliant LLMs for Global Language Environments

Speaker: Martin Jaggi (EPFL)

Date: Friday November 21, 2025 13:15-14:15 CET

Place: Physically at EPFL, BC Building, room BC 420

and online

<https://uzh.zoom.us/j/66608357823?pwd=cNAvZKibfpI06bb4zpgMKtREbngvZ.1>

Abstract

We present Apertus, a fully open suite of large language models (LLMs) designed to address two systemic shortcomings in today's open model ecosystem: data compliance and multilingual representation. Unlike many prior models that release weights without reproducible data pipelines or regard for content-owner rights, Apertus models are pretrained exclusively on openly available data, retroactively respecting robots.txt URL exclusions and filtering for non-permissive, toxic, and personally identifiable content. To mitigate risks of memorization, we adopt the Goldfish objective during pretraining, strongly suppressing verbatim recall of data while retaining downstream task performance. The Apertus models also expand multilingual coverage, training on 15T tokens from over 1800 languages, with ~40% of pretraining data allocated to non-English content. Released at 8B and 70B scales, Apertus approaches state-of-the-art results among fully open models on multilingual benchmarks, rivalling or surpassing open-weight counterparts. Beyond model weights, we release all scientific artifacts from our development cycle with a permissive license, including data preparation scripts, checkpoints, evaluation suites, and training code, enabling transparent audit and extension.

Link:

<https://www.swiss-ai.org/apertus>

Bio



MARTIN JAGGI is an associate professor at EPFL, where he is leading the Machine Learning and Optimization Laboratory.

Before that, he was a post-doctoral researcher at ETH Zurich, at the Simons Institute in Berkeley, and at École Polytechnique in Paris. He has earned his PhD in Machine Learning and Optimization from ETH Zurich in 2011, and a MSc in Mathematics also from ETH Zurich.